Is the impact of social distancing on coronavirus growth rates effective across different settings? A non-parametric and local regression approach to test and compare the 'doubling rate'.

Neil Lancastle

Faculty of Business and Law, De Montfort University

Correspondence: neil.lancastle@dmu.ac.uk

Abstract

Epidemiologists use mathematical models to predict epidemic trends, and these results are inherently uncertain when parameters are unknown or changing. In other contexts, such as climate, modellers use multi-model ensembles to inform their decision-making: when forecasts align, modellers can be more certain. This paper looks at a sub-set of alternative epidemiological models that focus on the 'doubling rate', and it cautions against relying on the method proposed in (Pike & Saini, 2020) which relies on the data for China to calculate future trajectories. Such approaches are subject to overfitting, a common problem in financial and economic modelling. This paper finds, surprisingly, that the data for China are hyper-exponential, not exponential. Instead, this paper proposes using non-parametric methods, and local regression methods, to support epidemiologists and policymakers in assessing the relative effectiveness of social distancing across multiple settings.

Acknowledgments

The author would like to thank Frank Kwabi, Dave Walsh, Tom Allen, Jonathan Rose, Linda Hickson, Olivia Lancastle, Fred Mear, Mark Kirstein and Stuart MacDonald for their useful comments.

Background

Mathematical models are routinely used in 'prediction of epidemic trends' (Pellis et al., 2020). The inherent uncertainty in these models, and in the early stages of an epidemic, is evident from the calls from epidemiologists at Oxford, UK, for 'large scale serological surveys' (Lourenço et al., 2020) to improve the accuracy of parameters such as susceptibility; and in the range of scenarios presented by epidemiologists at Imperial, UK (Ferguson et al., 2020). This difficult in choosing and parameterising the most effective prediction model is not unique to epidemiology. To resolve this, climate modellers use multi-model ensembles in their decision-making processes, acknowledging that 'even if the perfect climate model did exist, any projection is always conditional on the scenario considered'(Tebaldi & Knutti, 2007). Similarly, after the 2008 global financial crisis, over-reliance on equilibrium models by policymakers was recognised as both 'ubiquitous' (Bezemer, 2010) and 'flawed' (Bieta et al., 2012); leading to widespread calls for alternative approaches.

Motivation

All models need to meet some basic criteria: they need to be theoretically grounded, and they need to be replicable across different settings, preferably with out-of-sample back-testing. This paper is motivated by a plethora of articles that focus on the 'doubling rate', which is the time taken for the total number of people infected to double. A successful reduction in the doubling rate, by both individual and public health measures, is the goal of social distancing and other interventions. The aim is to flatten the curve and avoid a huge peak that would overwhelm health services (Wighton, 2020). The recent spate of articles includes contributions from non-epidemiologists and the media (Cuffe & Jeavans, 2020; Financial Times, 2020; Pike & Saini, 2020). The author is not arguing for alternative approaches to be given equal weight in the decision-making process, but for them to inform the debate as to whether alternative approaches are robust to out-of-sample back-testing. In addition, this research is motivated by a desire to test the validity of assumptions made by Pike and Saini (2020), who used data from China to predict out of sample death rates in other countries after the implementation of social distancing.

Ethics

There is a well-established principle in econometrics that using a limited dataset to forecast is generally frowned upon as overfitting (Bailey et al., 2016; Chalana, 2017). For that reason alone, the author would caution against relying too heavily on the data from China. Following a well-established precautionary principle that, in medical research, 'one should take reasonable measures to avoid threats that are serious and plausible' (Resnik, 2004), the author has refrained from using recent data on death statistics published by individual countries, and has instead focused on infection statistics (CSSE, 2020; UK-Government, 2020). Forecasting health care needs is known to be uncertain (Grenfell et al., 1994; Rein et al., 2011) but infection rates have the advantage that they are a leading indicator for pressure on intensive care beds. The author is required to get approval from his University ethics committee for research that 'would induce psychological stress, anxiety or humiliation' (DMU, 2020) and, given the author is not an epidemiologist, full approval and peer review were sought before pre-publication. Underestimates of intensive care beds could lead to a reduction in compliance with social distancing which, in turn, could lead increased harm to those most at risk of infection. Conversely, overestimates of intensive care beds could lead to unnecessary public health measures which, in turn, could lead to additional financial and economic harm.

Research question and data

Is the impact of social distancing on coronavirus growth rates comparable across different settings?

Following the precautionary principle above, the null hypothesis proposed is that social distancing in different contexts is likely to lead to different outcomes. This is to avoid a Type 1 error – the rejection of a true null hypothesis – the equivalent of convicting an innocent person in a criminal trial. Historical data of total infections in China are used. For other countries, data are compared at the relative lockdown dates shown in **Appendix 1**. The data were taken from Worldometers (Dadax, 2020) and links to the data for this paper are available in **Appendix 2**; and the R scripts used to run the tests are shown in **Appendix 3**. The author intends to add further countries, and to provide daily updates of **Figure 3** and **Figure 4**, via Twitter, until such time as the analysis can be provided by a third-party website. All works contained herein are provided free to use worldwide by the author under CC BY 2.0

Method

The first step in time series analysis is to collect and cleanse the data. This process requires transparency regarding how the data have been collected and handled, with any discrepancies resolved before analysis (DHSC, 2020). In finance and economics, the method most used is to winsorize to a stated range or percentage (Kroszner et al., 2007). Given the growth rate in total infections for China fell below 0.005% after 45 days, data after this date are excluded. The author excluded data from February 12th, when it was reported that infections in China jumped due to a change in the measurement method (Feuer, 2020) Model fitting using the full data is shown in **Appendix 4** as a robustness check.

An assumption that growth is exponential, in the early stages of an epidemic, is supported by the literature (Chowell et al., 2015; Nishiura et al., 2010), although the author is aware that other approximations exist, including polynomial. This paper does not attempt to review the theory behind this. For exponential-type growth, there are three choice of mathematical models: exponential-unrestrained; log-log (as per exponential, but with a longer period of decline); and logistic (approximately exponential at first, but with a reduced rate as an upper limit is approached). A logistic model fails the precautionary principle in that the carrying capacity, or c, can only be estimated ex-post. Hence, a logistic model is discarded.

The models being tested are therefore:

Model 1 – regression on ln(total cases) v time: $y = ae^{kt}$ *or* $ln(y) = a' + kt$

Model 2 - regression on ln(total cases) v ln(time): $\ln(y) = a + bln(t)$

Where y = total cases; t = time; and a, a' *and b are constants*

In financial time series analysis, for a stochastic (i.e.: random walk) process such as stock price movements, the rate of change in prices is generally used $\frac{p_t}{p_{t-1}}$: this technique is also applied here as $\Delta y = ln(y_t) - ln(y_{t-1}) \equiv \frac{y_t}{y_{t-1}}$, as a way to model when there are no new cases and the rate of change falls to zero ($\ln\left(\frac{y}{y}\right) = \ln(1) = 0$). Hence, two further models are tested, namely:

Model 3 – regression on daily change in total cases v time: $\frac{y_t}{y_{t-1}} = ae^{kt}$ *or* $\Delta y = a' + kt$

Model 4 - regression on daily change in total cases v ln(time): $\ln(y) = a + bln(t)$

Lastly, $\ln(\Delta y)$, a hyper-exponential model, is also tested[1]. In hyper-exponential decay, as the rate of change approaches zero ($y_t \rightarrow y_{t-1}$) then ($\ln\left(\frac{y_t}{y_t - 1}\right) \rightarrow -\infty$). For example, when the daily change in new cases falls to 1%, $\ln(\Delta y)$ is -4.6. Model 6 is included for completeness.

Model 5 – hyper-exponential: $\Delta y = ae^{kt}$ *or* $ln(\Delta y) = a' + kt$

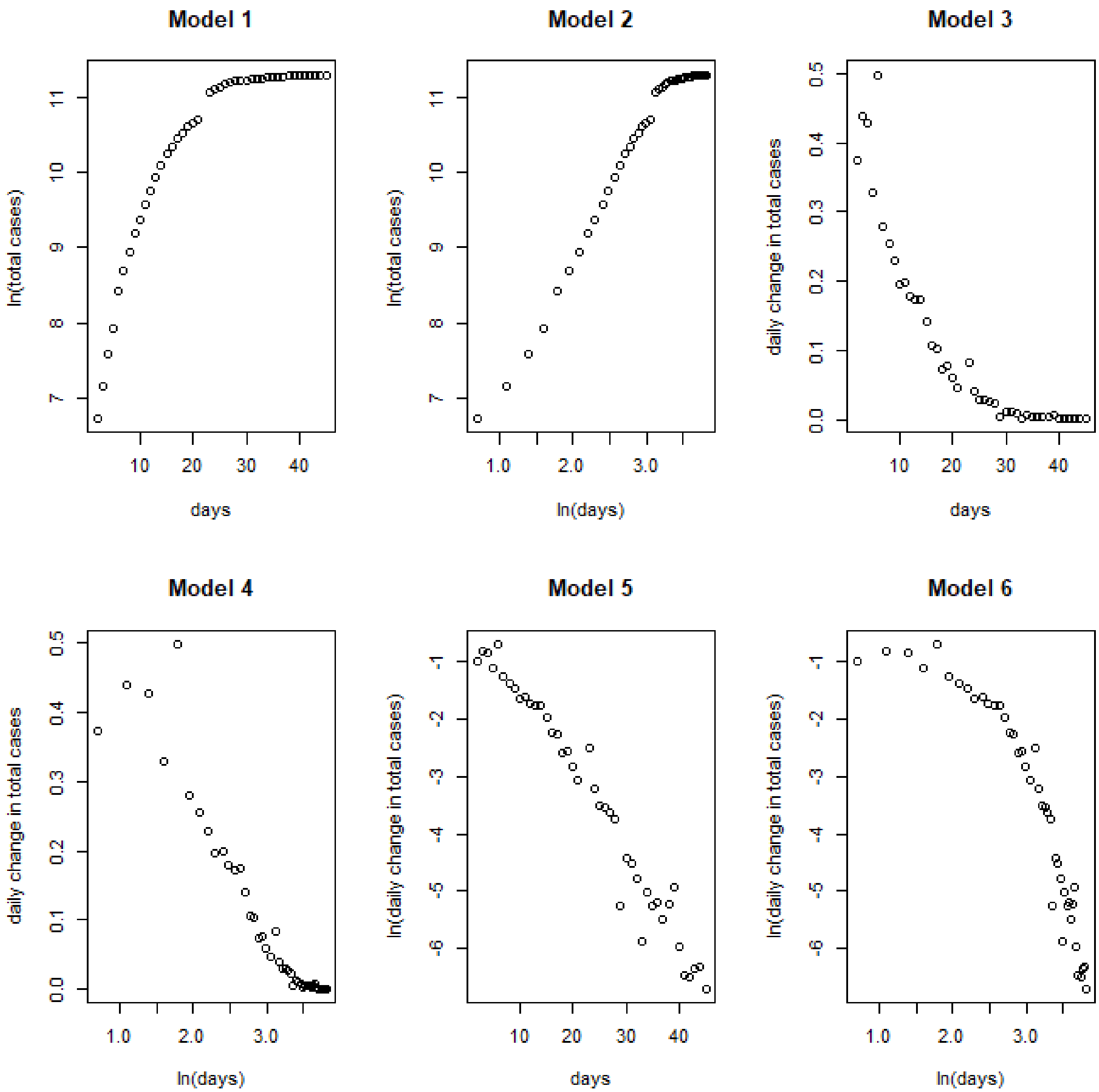
Model 6 – hyper-exponential with ln(time): $ln(\Delta y) = a + bln(t)$

The next step in time series forecasting, after data cleansing, is to analyse the goodness-of-fit to alternative models. This is done graphically, by plotting time (x axis) against each model – the closer the result is to a straight line, the better the goodness-of-fit.

Figure 1 plots the results for winsorized data from China. These confirm that Model 4 and Model 5 are good fits. The author understands that log(time) relationships are found in nature (Dahly, 2017) as are hyper-exponential (Varfolomeyev & Gurevich, 2001).

[1] The author understands this to be hyper-exponential or Laplace distribution; and is in discussion with colleagues to better improve his understanding of the mathematics and terminology for Models 3,4,5 and 6

Figure 1: Model fitting



We would not expect the data to be suitable for OLS regression, given there is convergence towards zero. This is confirmed by the descriptive statistics and Kolmogorov–Smirnov test in **Table 1**:

Table 1: Descriptive statistics

Time Series	Mean	Median	Skew	Kurtosis	KS Test
Total cases	49578	66492	-0.397	1.47	Reject
Daily change	0.109	0.041	1.35	3.79	Reject
ln(daily change)	-3.47	-3.21	-0.199	1.66	Reject

Therefore, not only should the reader be wary of overfitting using the China data to extrapolate; the reader should also be wary of overfitting via OLS (linear) regression.

The next part of this analysis applies a non-parametric method to determine whether an out-of-sample forecast, from observations that were not part of the sample, is possible. From Kruskal-Wallis tests on the daily changes for China (not shown) – splitting the data into the three periods (period 1: 0-9 days; period 2: 10-19 days; and period 3: 20-20 days) – the rate of change was found to be significantly different. To illustrate this, **Appendix 5** and **Appendix 6** show locally estimated scatterplot smoothing (LOESS) using Model 4 and Model 5 respectively. The three periods are clearly different in Model 4, which is the second differential used for fitting in the paper by (Pike & Saini, 2020). Model 5 is more linear, but with the caveats that i) the model is only as good as the data and ii) there is no theory, to the author's knowledge, that supports the idea of hyper-exponential delay and iii) the author is not an epidemiologist. In both models, the results show a significant drop in infections over three distinct phases after social distancing in China; and the results provide researchers with some visible patterns that might, or might not, emerge from the data for other countries as social distancing continues.

The last part of this analysis uses Model 4 and Model 5 to compare six countries - China, Italy, Spain, France, US and UK – using LOESS. In Model 4, when there are no new daily cases the rate of change falls to zero. In Model 5, when the number of new cases approaches zero, the model approaches $-\infty$; alternatively, when the number of new cases falls to 1%, the model is approximately -4.6. **Table 2** shows how to interpret the X-axis and Y-axis for both models; **Figure 2** and **Figure 3** show results graphically; and Table 3 summarises paModel mcurrent values as at i) 10 days after lockdown and ii) 31st March 2020:

Table 2: Interpreting Model 4 and Model 5

Day X-axis	Total Cases	% daily increase	Model 4 Y-axis	Model 5 Y-axis	Growth rate
0	100				
1	200	100%	0.69	-0.37	Doubling every day
2	400	100%	0.69	-0.37	
3	600	50%	0.41	-0.90	Doubling every two days
4	800	33%	0.29	-1.25	Doubling every three days
5	1000	25%	0.22	-1.50	Doubling every four days
6	1010	1%	0.01	-4.61	1% growth per day

Figure 2 – Model 4 - Comparative Growth Rates for China, Italy, Spain, France, US and UK

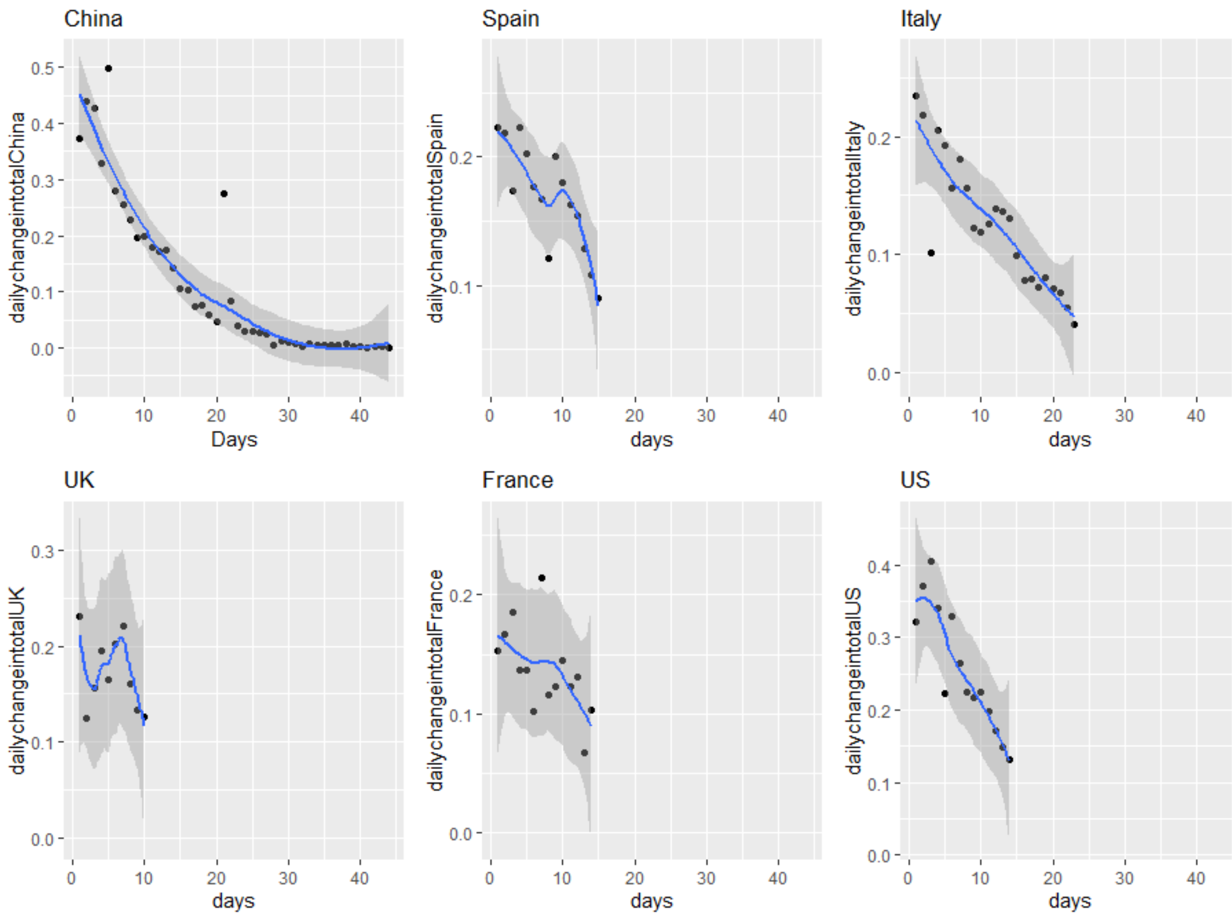


Figure 3 – Model 5 - Comparative Growth Rates for China, Italy, Spain, France, US and UK

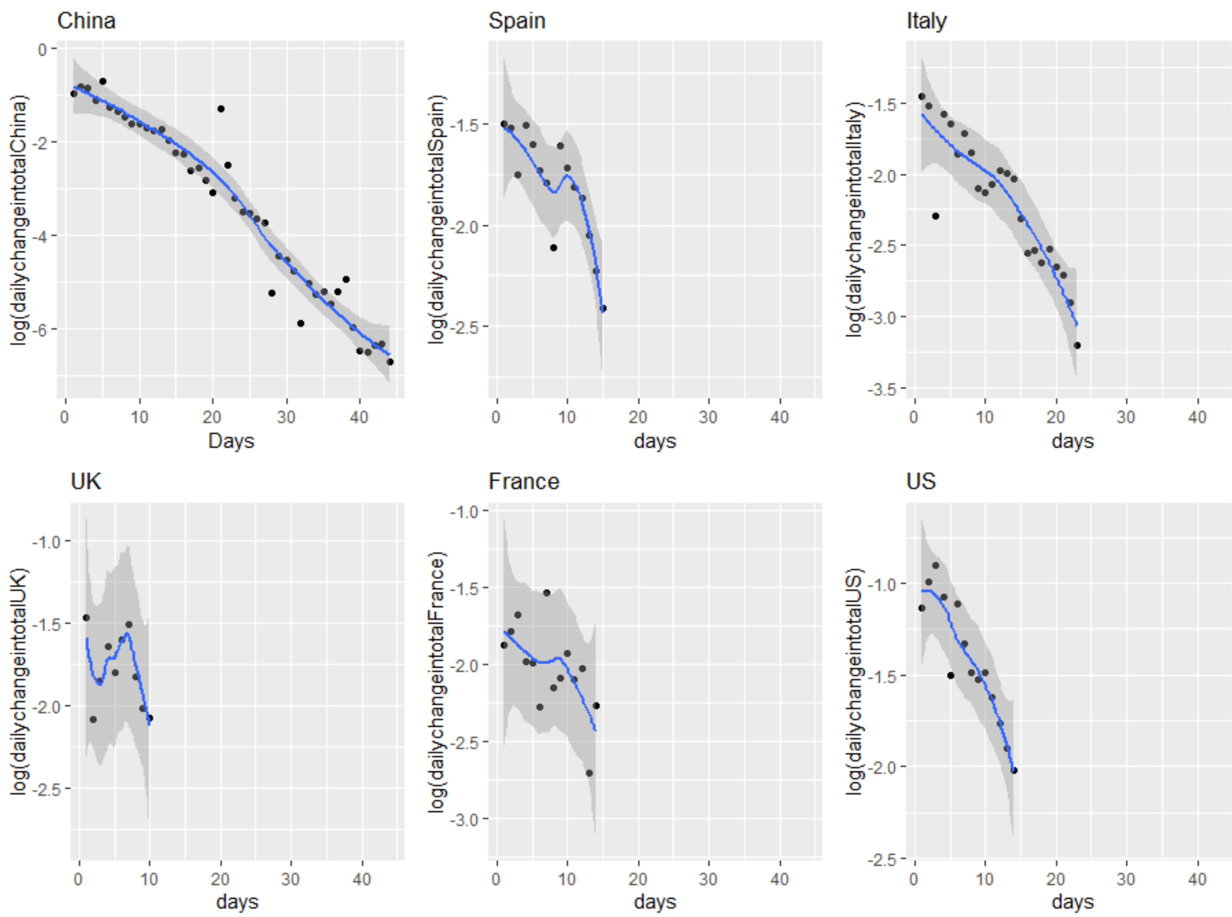


Table 3: Country Comparisons as at 31st March 2020

	10 days after lockdown		31st March 2020		
Country	**Model 4 Y-axis**	**Model 5 Y-axis**	**Model 4 Y-axis**	**Model 5 Y-axis**	**Current daily growth rate**
US	0.2252	-1.49	0.1326	-2.02	14%
UK	0.1259	-2.07	0.1259	-2.07	13%
France	0.1445	-1.93	0.1034	-2.27	11%
Spain	0.1798	-1.72	0.0897	-2.41	9%
Italy	0.1187	-2.13	0.0406	-3.2	4%
China	0.1985	-1.62	0.0012	-6.73	0.01%

Conclusion

The methodology presented here is a non-parametric and graphical approach, using freely available software, to test whether different phases of social distancing are comparable across different settings. If the null hypothesis, that social distancing is not comparable, can be rejected, then policymakers can be reassured that their social distancing measures are being successful.

With the above caveats above data quality, transparency, and consistency, the data for China show that strong social distancing led initially to a high reduction in the doubling rate; then a rapid decline towards a daily doubling rate of <1% by 30 days; followed by a stable period of low transmission. There is some evidence, from the data shown in Figures 2 and 3, that other countries in the panel have shown similar trajectories since social distancing.

This paper cautions against policy based on the second derivative method proposed in (Pike & Saini, 2020) for three main reasons: first, to the author's knowledge, there is no theoretical basis for using a second derivative (Method 4); second, the China data fit a hyper-exponential model better (Model 5) and this is particularly the case in period 3, 20-29 days, of social distancing (compare Appendix 5 and Appendix 6); third, there are likely to be very different data collection, quality, and consistency issues in different settings.

These are tentative conclusions, and the results require replicated and confirmation – this is a pre-print of the paper. With these caveats; this approach, with the addition of further countries by the author and other researchers, could inform the policy debate around the effectiveness of different approaches of social distancing. There are three parts to the methodology: i) non-parametric tests, to compare Model 4 and Model 5 parameters during each phase of social distancing ii) a graphical approach, to determine whether the data in different contexts follow the same trajectories under both Model 4 and Model 5 and iii) LOESS to check for similar trajectories in equivalent periods: 0 days; 10-19 days; and 20-29 days.

In terms of prediction, the author supports those epidemiologists who are calling for better data (Lourenço et al., 2020) through testing, and who are presenting a range of scenarios (Ferguson et al., 2020); until such time that a consensus emerges from a multi-modelling perspective. In particular, the author cautions against giving much weight to models that extrapolate from growth rates, or that assume other countries will necessarily follow the same path as China, in order to avoid a type 1 error and rejection of a true null hypothesis. To paraphrase the infinite monkey theorem, that is not to say that, with enough predictive models, at least one of them will fit the available data *by chance*. Instead, following the precautionary principle, the author recommends against policy recommendations that are not supported by clinical observations in the field and/or by a convergence of predictions from different epidemiological prediction models, such as the SIR models at Oxford, UK and Imperial, UK.

Lastly, as well as using this method to compare different phases of social distancing across different settings, the methodology might also be useful, *ex-post*, for sociologists, economists and political scientists to determine the social, economic, and political costs of different approaches to social distancing in different settings.

References

Bailey, D. H., Zhu, Q., & Marcos, L. (2016). *Backtest overfitting in financial markets*. 1–8.

Bezemer, D. (2010). Understanding financial crisis through accounting models. *Accounting, Organizations and Society*, *35*(7), 676–688. https://doi.org/10.1016/j.aos.2010.07.002

Bieta, V., Milde, H., & Weber, N. M. (2012). "A Flaw in the Model … that Defines How the World Works." *International Journal of Financial Research*, *3*(1). https://doi.org/10.5430/ijfr.v3n1p81

Chalana, A. (2017). *Predicting and preventing overfitting of financial models*. 1–13.

Chowell, G., Viboud, C., Hyman, J., & Simonsen, L. (2015). The Western Africa Ebola Virus Disease Epidemic Exhibits Both Global Exponential and Local Polynomial Growth Rates. *Plos Currents*, *21*(7). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4322058/

CSSE. (2020). *Coronavirus COVID-19 Global Cases*. https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6

Cuffe, R., & Jeavans, C. (2020, March 28). How the UK's coronavirus epidemic compares to other countries. *BBC News Online*. https://www.bbc.co.uk/news/health-52056111

Dadax. (2020). *Worldometers*. https://www.worldometers.info/coronavirus/#countries

Dahly, D. (2017). *Log Log Regression*. https://darrendahly.github.io/post/loglog/

DHSC. (2020). *Update on our reporting process*. https://twitter.com/DHSCgovuk/status/1243237211119800323

DMU. (2020). *Research requiring ethical approval*. https://www.dmu.ac.uk/research/ethics-and-governance/research-requiring-ethical-approval.aspx

Ferguson, N. M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G., Dighe, A., Dorigatti, I., Fu, H., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Okell, L. C., van Elsland, S., … Ghani, A. C. (2020). *Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand*. *March*, 20. https://doi.org/10.25561/77482

Feuer, W. (2020, February 26). 'Confusion breeds distrust:' China keeps changing how it counts coronavirus cases. *CNBC*. https://www.cnbc.com/2020/02/26/confusion-breeds-distrust-china-keeps-changing-how-it-counts-coronavirus-cases.html

Financial Times. (2020). *Coronavirus tracked: the latest figures as the pandemic spreads*. https://www.ft.com/coronavirus-latest

Grenfell, B. T., Kleczkowski, A., Ellner, S. P., & Bolker, B. M. (1994). Measles as a case study in nonlinear forecasting and chaos. *Philosophical Transactions - Royal Society of London, A*, *348*(1688), 497–514. https://doi.org/10.1098/rsta.1994.0108

Kroszner, R. S., Laeven, L., & Klingebiel, D. (2007). Banking crises, financial dependence, and growth. *Journal of Financial Economics*, *84*(1), 187–228. https://doi.org/10.1016/j.jfineco.2006.05.001

Lourenço, J., Paton, R., Ghafari, M., Kraemer, M., Thompson, C., Klenerman, P., & Gupta, S. (2020). *Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic*. 1–7.

Nishiura, H., Chowell, G., Safan, M., & Castillo-Chavez, C. (2010). Pros and cons of estimating the reproduction number from early epidemic growth rate of influenza A (H1N1) 2009. *Theoretical Biology and Medical Modelling*, *7*(1), 1–13. https://doi.org/10.1186/1742-4682-7-1

Pellis, L., Cauchemez, S., Ferguson, N. M., & Fraser, C. (2020). Systematic selection between age and household structure for models aimed at emerging epidemic predictions. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-019-14229-4

Pike, W. T., & Saini, V. (2020). *An international comparison of the second derivative of COVID-19 deaths after implementation of social distancing measures*. 1–7.

Rein, D. B., Wittenborn, J. S., Weinbaum, C. M., Sabin, M., Smith, B. D., & Lesesne, S. B. (2011). Forecasting the morbidity and mortality associated with prevalent cases of pre-cirrhotic chronic hepatitis C in the United States. *Digestive and Liver Disease*, *43*(1), 66–72. https://doi.org/10.1016/j.dld.2010.05.006

Resnik, D. B. (2004). The precautionary principle and medical decision making. *Journal of Medicine and Philosophy*, *29*(3), 281–299. https://doi.org/10.1080/03605310490500509

Tebaldi, C., & Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *365*(1857), 2053–2075. https://doi.org/10.1098/rsta.2007.2076

UK-Government. (2020). *COVID-19: track coronavirus cases*. https://www.gov.uk/government/publications/covid-19-track-coronavirus-cases

Varfolomeyev, S. D., & Gurevich, K. G. (2001). The hyperexponential growth of the human population on a macrohistorical scale. *Journal of Theoretical Biology*, *212*(3), 367–372. https://doi.org/10.1006/jtbi.2001.2384

Wighton, K. (2020). *Individual response to COVID-19 'as important' as government action*. https://www.imperial.ac.uk/news/195976/individual-response-covid-19-important-government-action/

Appendix 1

first day of lockdowns – day 0

China 23rd January (24th January - most Wuhan provinces) - 830 cases

UK Saturday 21st March (announced 5pm on Friday 20th but pubs still open) - 5018 cases

Spain Sunday 15th March (announced on Saturday evening) - 7375 cases

Italy Sunday 8th March (announced early in morning) - 7375 cases

France Tuesday 17th March (announced on Monday to start next day) - 7730 cases

US Tuesday 16th March (announced on Monday to start next day) - 4596 cases

Appendix 2 – Data

Available here: https://drive.google.com/drive/folders/1oEXaapNeVBGGvzOTvmweie64r8JSzH-j?usp=sharing

Appendix 3 – R Scripts

```
install.packages("MASS")
library(MASS)

install.packages("moments")
library(moments)

install.packages("ggplot2")
library("ggplot2")

# avoid spurious accuracy
op <- options(digits = 3)

# Read in csv file
```

```
china <- read.table("c:/data/china.csv", header = TRUE, sep = ",")
chinawinz <- read.table("c:/data/chinawinz.csv", header = TRUE, sep = ",")
chinaranksum <- read.table("c:/data/chinaranksum.csv", header = TRUE, sep = ",")


# FIGURE 1 Model fitting

par(mfrow=c(2,3))

plot(china$Days, log(china$COVID),
   xlab = "days", ylab="ln(total cases)",
   main = "Model 1")

plot(log(china$Days), log(china$COVID),
   xlab = "ln(days)", ylab="ln(total cases)",
   main = "Model 2")

plot(china$Days, china$Diff,
   xlab = "days", ylab="daily change in total cases",
   main = "Model 3")

plot(log(china$Days), china$Diff,
   xlab = "ln(days)", ylab="daily change in total cases",
   main = "Model 4")

plot(china$Days, log(china$Diff),
   xlab = "days", ylab="ln(daily change in total cases)",
   main = "Model 5")
```

```
plot(log(china$Days), log(china$Diff),
    xlab = "ln(days)", ylab="ln(daily change in total cases)",
    main = "Model 6")

# FIGURE 1 Model fitting - excluding February 12th

par(mfrow=c(2,3))

par(mfrow=c(2,3))

plot(chinawinz$Days, log(chinawinz$COVID),
    xlab = "days", ylab="ln(total cases)",
    main = "Model 1")

plot(log(chinawinz$Days), log(chinawinz$COVID),
    xlab = "ln(days)", ylab="ln(total cases)",
    main = "Model 2")

plot(chinawinz$Days, chinawinz$Diff,
    xlab = "days", ylab="daily change in total cases",
    main = "Model 3")

plot(log(chinawinz$Days), chinawinz$Diff,
    xlab = "ln(days)", ylab="daily change in total cases",
    main = "Model 4")

plot(chinawinz$Days, log(chinawinz$Diff),
    xlab = "days", ylab="ln(daily change in total cases)",
    main = "Model 5")
```

```
plot(log(chinawinz$Days), log(chinawinz$Diff),
    xlab = "ln(days)", ylab="ln(daily change in total cases)",
    main = "Model 6")



# Table 1 Descriptive statistics

# Model 1 - ln(total cases)
summary(chinawinz$COVID)
skewness(chinawinz$COVID)
kurtosis(chinawinz$COVID)

x <- rnorm(1000)
# Do x and y come from the same distribution?
ks.test(chinawinz$COVID, x)

Model 3 - daily change in total cases
summary(chinawinz$Diff)
skewness(chinawinz$Diff)
kurtosis(chinawinz$Diff)

# Do x and y come from the same distribution?
ks.test(chinawinz$Diff, x)

Model 5 - ln(daily change in total cases)
summary(chinawinz$Diff)
skewness(chinawinz$Diff)
```

```
kurtosis(chinawinz$Diff)

# Do x and y come from the same distribution?
ks.test(log(chinawinz$Diff), x)
summary(log(chinawinz$Diff))
skewness(log(chinawinz$Diff))
kurtosis(log(chinawinz$Diff))

# Appendix 5 Model 4 LOESS for China periods 0-9, 10-19 and 20-29 days

install.packages("gridExtra")
library("gridExtra")

three <- read.table("c:/data/three.csv", header = TRUE, sep = ",")

attach(three)
period1 = Days1
period2 = Days2
period3 = Days3
dailychangeintotalcases1 = Diff1
dailychangeintotalcases2 = Diff2
dailychangeintotalcases3 = Diff3

p1 <- ggplot(three, aes(period1, dailychangeintotalcases1)) +
  geom_point()

# loess method: local regression fitting
p1 <- p1 + geom_smooth(method = "loess", level = 0.99)
```

```
p2 <- ggplot(three, aes(period2, dailychangeintotalcases2)) +
  geom_point()

# loess method: local regression fitting
p2 <- p2 + geom_smooth(method = "loess", level = 0.99)

p3 <- ggplot(three, aes(period3, dailychangeintotalcases3)) +
  geom_point()

# loess method: local regression fitting
p3 <- p3 + geom_smooth(method = "loess", level = 0.99)

grid.arrange(p1, p2, p3, ncol=3)



# Appendix 6 Model 5 LOESS for China periods 0-9, 10-19 and 20-29 days

install.packages("gridExtra")
library("gridExtra")

three <- read.table("c:/data/three.csv", header = TRUE, sep = ",")

attach(three)
period1 = Days1
period2 = Days2
period3 = Days3
ln_dailychangeintotalcases1 = log(dailychangeintotalcases1)
ln_dailychangeintotalcases2 = log(dailychangeintotalcases2)
```

```
ln_dailychangeintotalcases3 = log(dailychangeintotalcases3)

p1 <- ggplot(three, aes(period1, ln_dailychangeintotalcases1)) +
  geom_point()

# loess method: local regression fitting
p1 <- p1 + geom_smooth(method = "loess", level = 0.99)


p2 <- ggplot(three, aes(period2, ln_dailychangeintotalcases2)) +
  geom_point()

# loess method: local regression fitting
p2 <- p2 + geom_smooth(method = "loess", level = 0.99)

p3 <- ggplot(three, aes(period3, ln_dailychangeintotalcases3)) +
  geom_point()

# loess method: local regression fitting
p3 <- p3 + geom_smooth(method = "loess", level = 0.99)

grid.arrange(p1, p2, p3, ncol=3)


# Figures 2 and 3

# Read in csv file
alldata <- read.table("c:/data/alldata.csv", header = TRUE, sep = ",")
```

```
attach(alldata)
days = Days

# Model 4

plot1  <- ggplot(alldata, aes(Days, dailychangeintotalChina)) +
  geom_point()+ ggtitle("China")

# loess method: local regression fitting
p1 <- plot1 + geom_smooth(method = "loess", level = 0.99)



plot2 <- ggplot(alldata, aes(days, dailychangeintotalUK)) +
  geom_point()+ ggtitle("UK")

# loess method: local regression fitting
p2 <- plot2 + geom_smooth(method = "loess", level = 0.99)

plot3 <- ggplot(alldata, aes(days, dailychangeintotalSpain)) +
  geom_point()+ ggtitle("Spain")

# loess method: local regression fitting
p3 <- plot3 + geom_smooth(method = "loess", level = 0.99)

plot4 <- ggplot(alldata, aes(days, dailychangeintotalItaly)) +
  geom_point()+ ggtitle("Italy")

# loess method: local regression fitting
p4 <- plot4 + geom_smooth(method = "loess", level = 0.99)
```

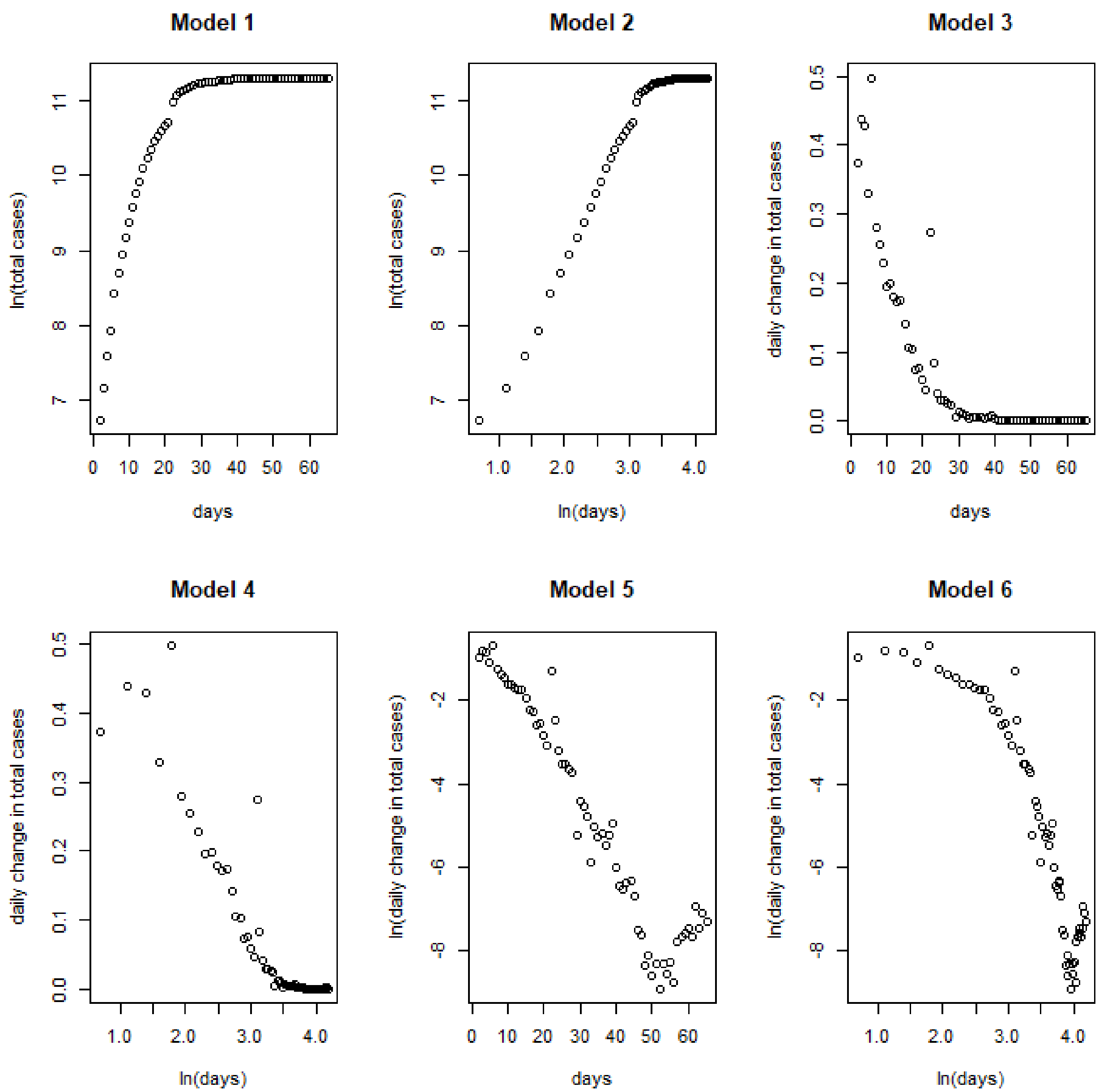
```
plot5 <- ggplot(alldata, aes(days, dailychangeintotalFrance)) +
  geom_point()+ ggtitle("France")

# loess method: local regression fitting
p5 <- plot5 + geom_smooth(method = "loess", level = 0.99)

plot6 <- ggplot(alldata, aes(days, dailychangeintotalUS)) +
  geom_point()+ ggtitle("US")

# loess method: local regression fitting
p6 <- plot6 + geom_smooth(method = "loess", level = 0.99)

grid.arrange(p1, p3, p4, p2, p5, p6, ncol=3)

# Model 5

log(dailychangeintotalChina)
log(dailychangeintotalUK)
log(dailychangeintotalSpain)

plot1  <- ggplot(alldata, aes(Days, log(dailychangeintotalChina))) +
  geom_point() + ggtitle("China")



# loess method: local regression fitting
p1 <- plot1 + geom_smooth(method = "loess", level = 0.99)
```

```
plot2 <- ggplot(alldata, aes(days, log(dailychangeintotalUK))) +
  geom_point()+ ggtitle("UK")


# loess method: local regression fitting
p2 <- plot2 + geom_smooth(method = "loess", level = 0.99)

plot3 <- ggplot(alldata, aes(days, log(dailychangeintotalSpain))) +
  geom_point()+ ggtitle("Spain")


# loess method: local regression fitting
p3 <- plot3 + geom_smooth(method = "loess", level = 0.99)

plot4 <- ggplot(alldata, aes(days, log(dailychangeintotalItaly))) +
  geom_point()+ ggtitle("Italy")


# loess method: local regression fitting
p4 <- plot4 + geom_smooth(method = "loess", level = 0.99)

plot5 <- ggplot(alldata, aes(days, log(dailychangeintotalFrance))) +
  geom_point()+ ggtitle("France")


# loess method: local regression fitting
p5 <- plot5 + geom_smooth(method = "loess", level = 0.99)

plot6 <- ggplot(alldata, aes(days, log(dailychangeintotalUS))) +
```

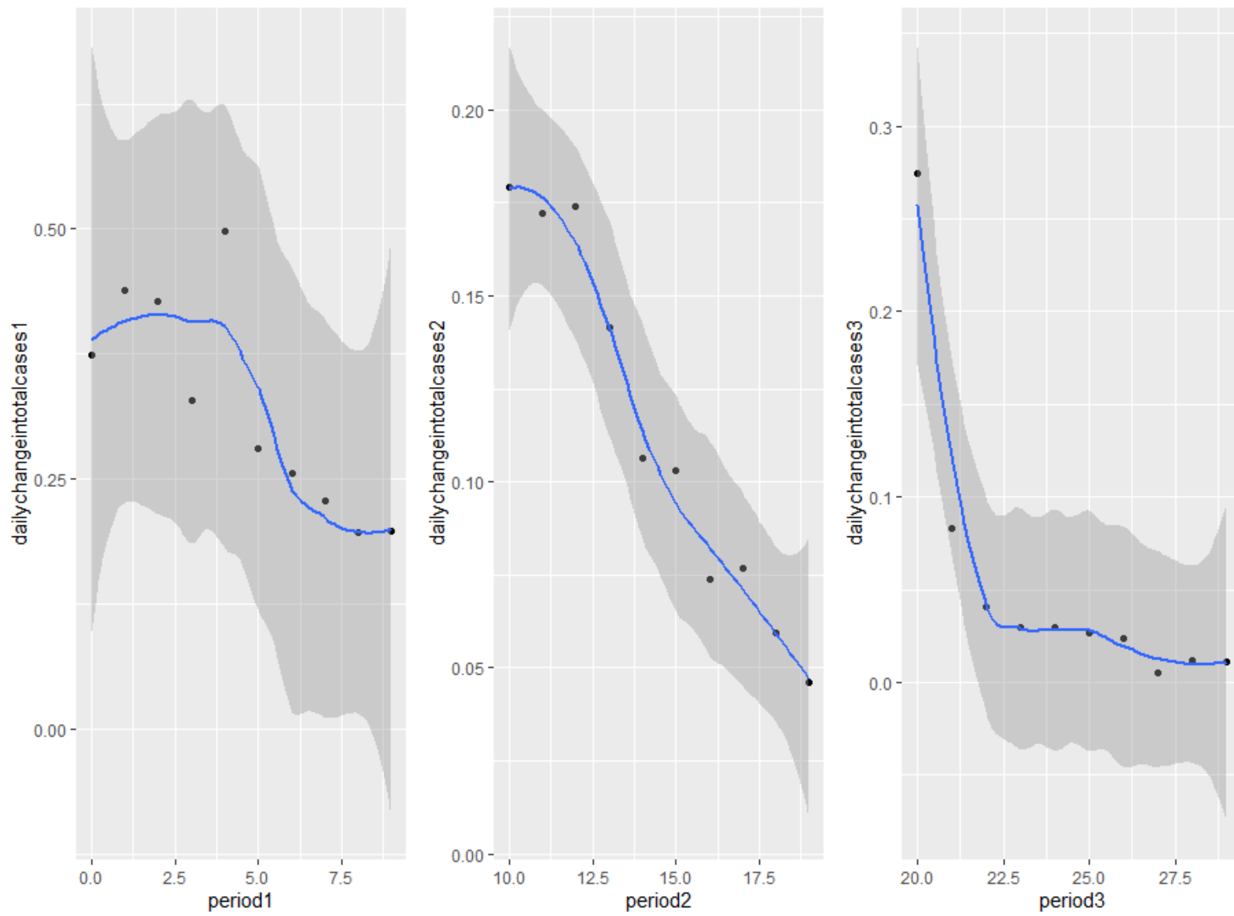
```
  geom_point()+ ggtitle("US")


# loess method: local regression fitting
p6 <- plot6 + geom_smooth(method = "loess", level = 0.99)

grid.arrange(p1, p3, p4, p2, p5, p6, ncol=3)

# announcement of lockdown
# China 23rd January (24th January - most Wuhan provinces) - 830 cases
# UK Saturday 21st March (announced 5pm on Friday 20th but pubs still open) - 5018 cases
# Spain Sunday 15th March (announced on Saturday evening) - 7375 cases
# Italy Sunday 8th March (announced early in morning) - 7375 cases
# France Tuesday 17th March (announced on Monday to start next day) - 7730 cases
# US Monday 16th March (announced on Monday to start next day) - 4596 cases
```

Appendix 4

Figure 5: Model fitting – full data



Appendix 5 – Model 4 - LOESS for China periods 0-9, 10-19 and 20-29 days



Appendix 6 – Model 5 - LOESS for China periods 0-9, 10-19 and 20-29 days